Monitoring for the Datalake

Rizart Dona

CERN

December 9, 2020 - 2nd ESCAPE WP2/DIOS Workshop

Overview

- Monitoring Stack

- Monitoring Architecture

- GFAL Dashboards

- FTS Dashboards

- Rucio Dashboards

- Future Work

- References
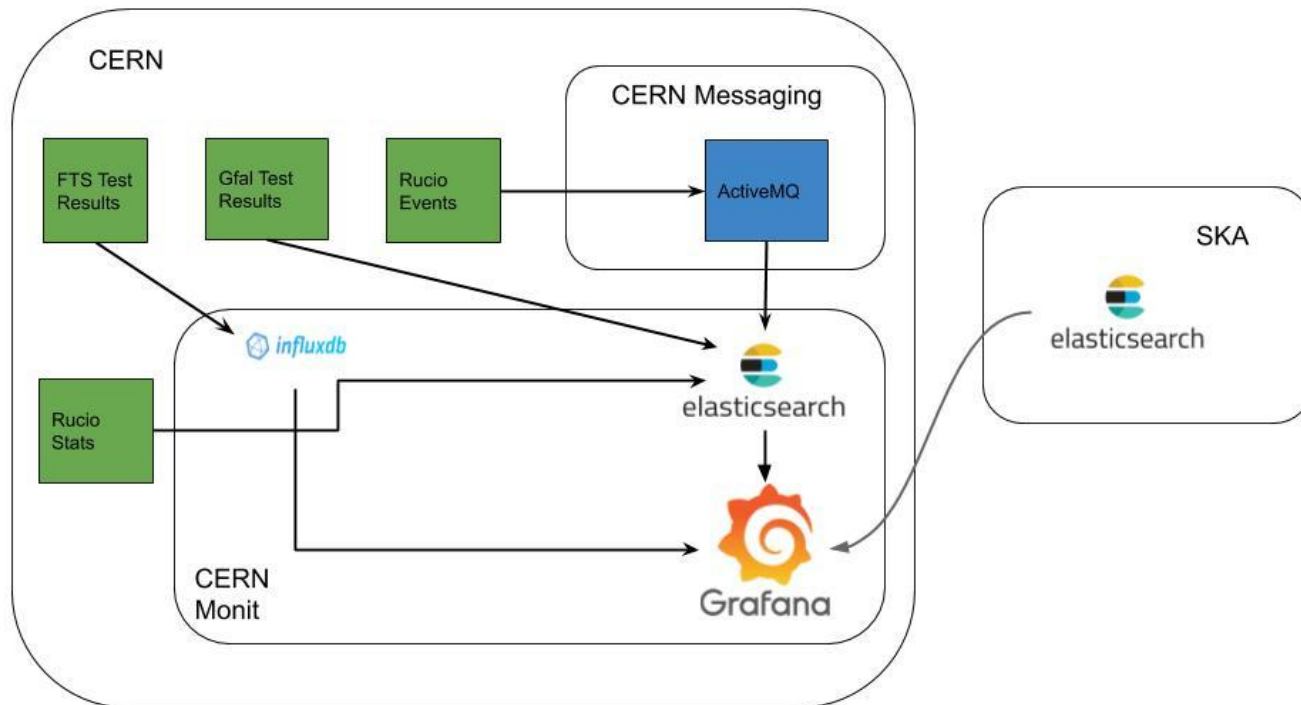
rizart.dona@cern.ch

Monitoring Stack

- Backend storage/messaging technologies used

  - **Elasticsearch**, a distributed multitenant-capable full-text search engine

  - **InfluxDB**, an open source time series database that can handle massive volumes and countless sources of time-stamped data

  - **ActiveMQ**, an open source, multi-protocol, Java-based messaging broker

- Visualization platform → **Grafana**

  - A multi-platform open source **analytics** and interactive **visualization** web application

  - Hosted @ CERN by Monit, separate ESCAPE organization

  - Granular permission schema, all interested users already registered

  - Supports multiple datasources (ES, InfluxDB, Graphite, MySQL, etc.)

- Our main datasources

  - **Elasticsearch** @ CERN (hosted by Monit) → Retention policy of 12 months

  - **InfluxDB** @ CERN → Retention policy of 1 month

  - **Elasticsearch** @ SKAO → No retention policy in place, limited by storage given

rizart.dona@cern.ch
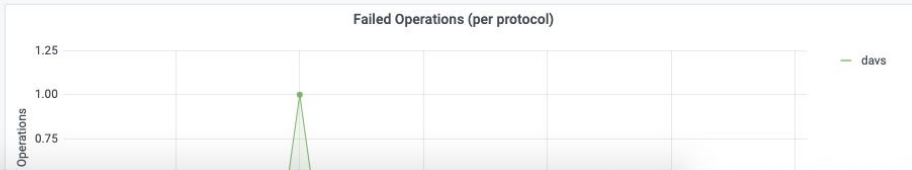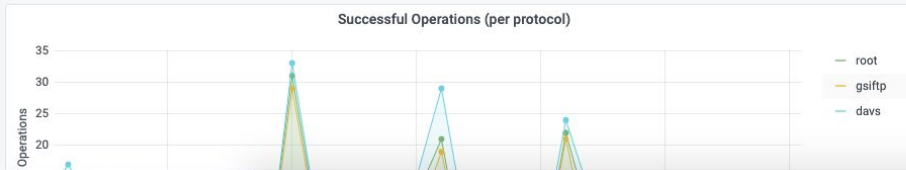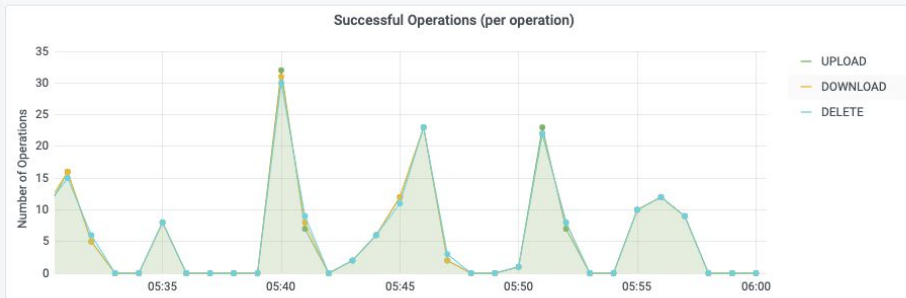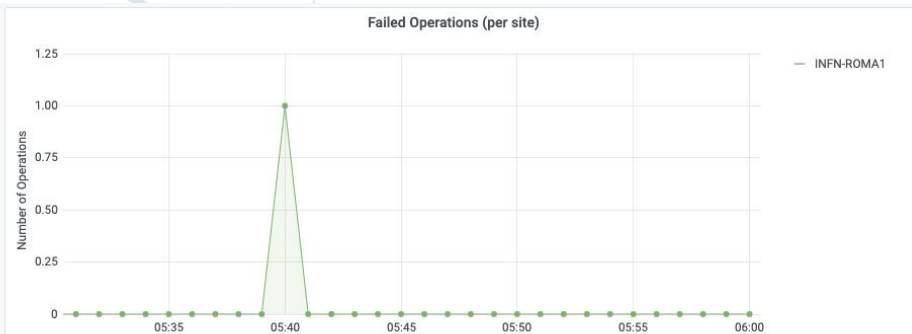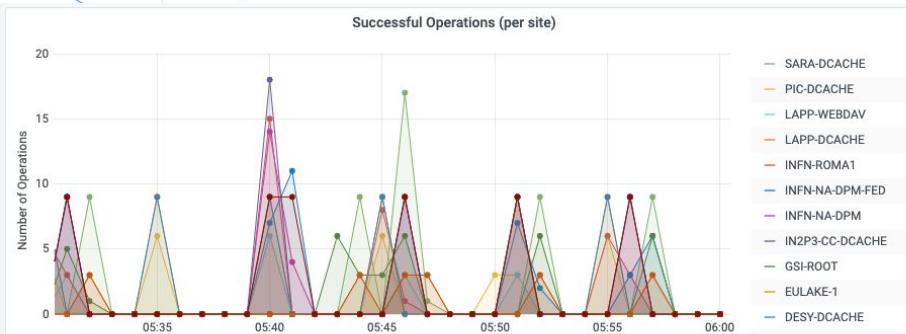
rizart.dona@cern.ch

GFAL Dashboards (1/2)

- Data produced and pushed by custom [python script](#)

- Basic GFAL upload/download/delete operations per endpoint

- A user can filter plots by

    - Endpoints

    - Operations (upload, download, delete) - SUCCEEDED/FAILED/SKIPPED

    - Protocol (gsiftp, root, http)

- ES datasource

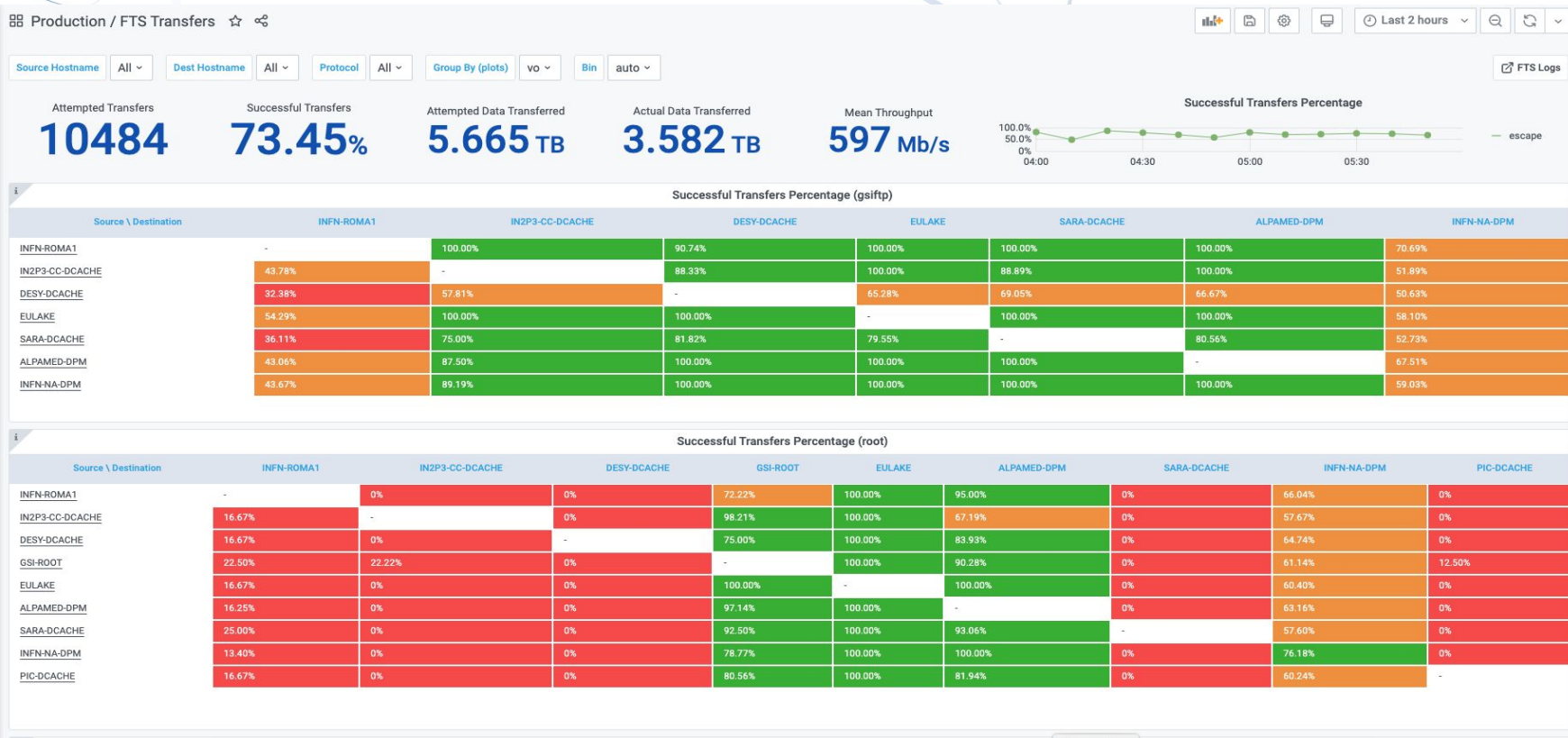rizart.dona@cern.ch

FTS Dashboards (1/3)

- Data produced by the fts-analysis-datalake toolkit, monitoring basically TPC transfers between the endpoints

- Automatically pushed by the FTS server to ES (configured by FTS team + Monit @ CERN)

- Both InfluxDB + ES datasources

  - Only InfluxDB currently used in the dashboards

  - InfluxDB going to be decommissioned soon → **Migration** of dashboards to ES based needed

- Main highlights for user

  - Aggregated stats (total data transferred, mean throughput, successful transfers, etc.)

  - FTS transfers **efficiency** matrix per protocol

  - Error codes & logfile links

- A lot of tweaking required on the creation aspect to get "beautiful" and intuitive panels

  - e.g. the mapping between endpoints → RSE names was done on the Grafana level

rizart.dona@cern.ch

FTS Dashboards (2/3)

rizart.dona@cern.ch

Funded by the European Union's
Horizon 2020 - Grant N° 824064

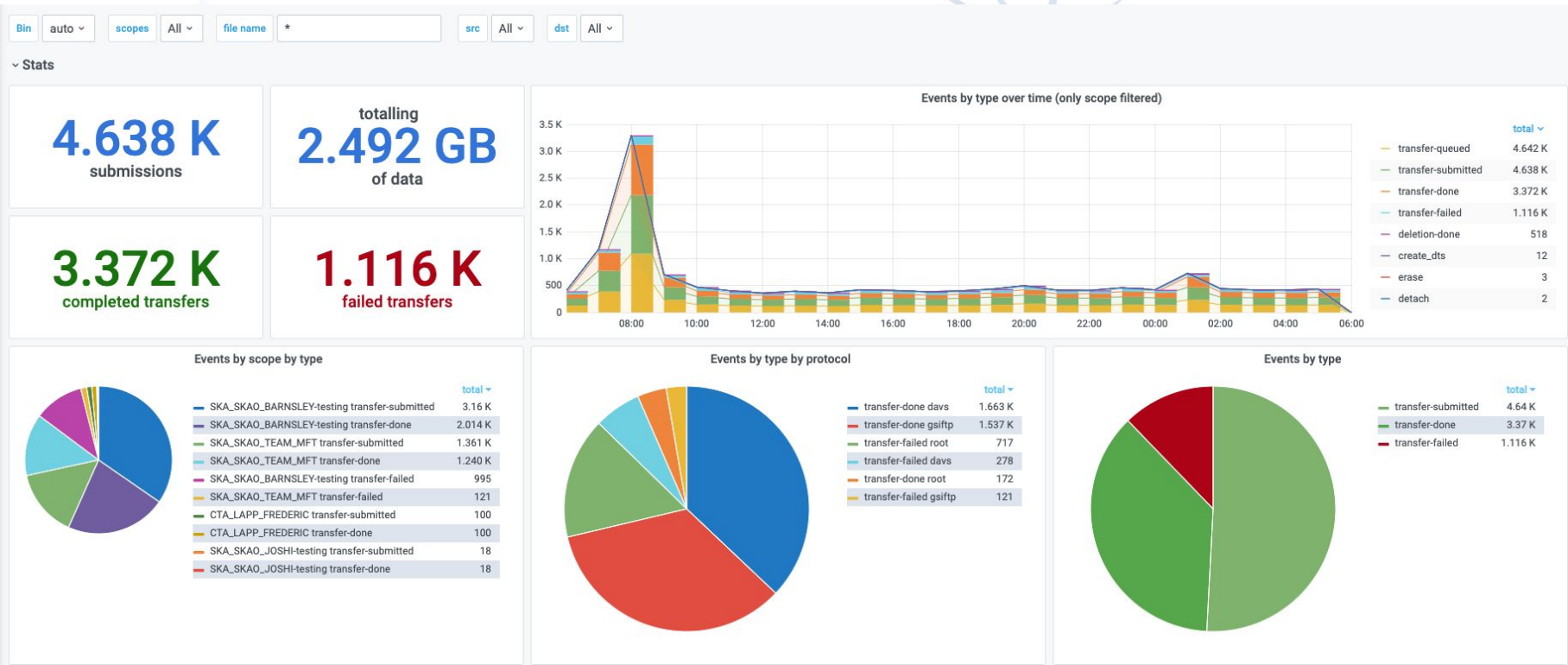# FTS Dashboards (3/3)

rizart.dona@cern.ch

Rucio Dashboards - Rucio Events (1/3)

- Data produced by a Rucio daemon (**Hermes**) → ActiveMQ broker → ES

- Monitoring **replica creation/deletion** (submissions, data transferred, failed transfers, creation of datasets, etc.)

- **Transfer matrix** helps user understand the connectivity between RSES

  - Percentages of successful replication (which are basically reflect into FTS TPC transfers)

  - Recently **optimized**, it feels snappier now (many thanks to Rob Barnsley for helping in this)

- Table with error codes and logfile links to the actual FST transfers

Rucio Dashboards - Rucio Events (2/3)

rizart.dona@cern.ch

rizart.dona@cern.ch

Funded by the European Union's
Horizon 2020 - Grant N° 824064

Rucio Dashboards - Rucio Statistics (1/3)

- Data source → cron job that fetches periodically (every 10 minutes) the relevant info, pushing to ES

- Novel dashboard, was extensively used during the **FDR** to monitor the progress through the day

- RSE stats
  - Used storage / RSE
  - Number of files / RSE

- Experiment stats
  - Used storage / Experiment
  - DIDs (files, datasets, containers) / Experiment

- Scope stats
  - Used storage / Scope
  - DIDs (files, datasets, containers) / Scope

Rucio Dashboards - Rucio Statistics (2/3)

rizart.dona@cern.ch

Rucio Dashboards - Rucio Statistics (3/3)

December 9, 2020

rizart.dona@cern.ch
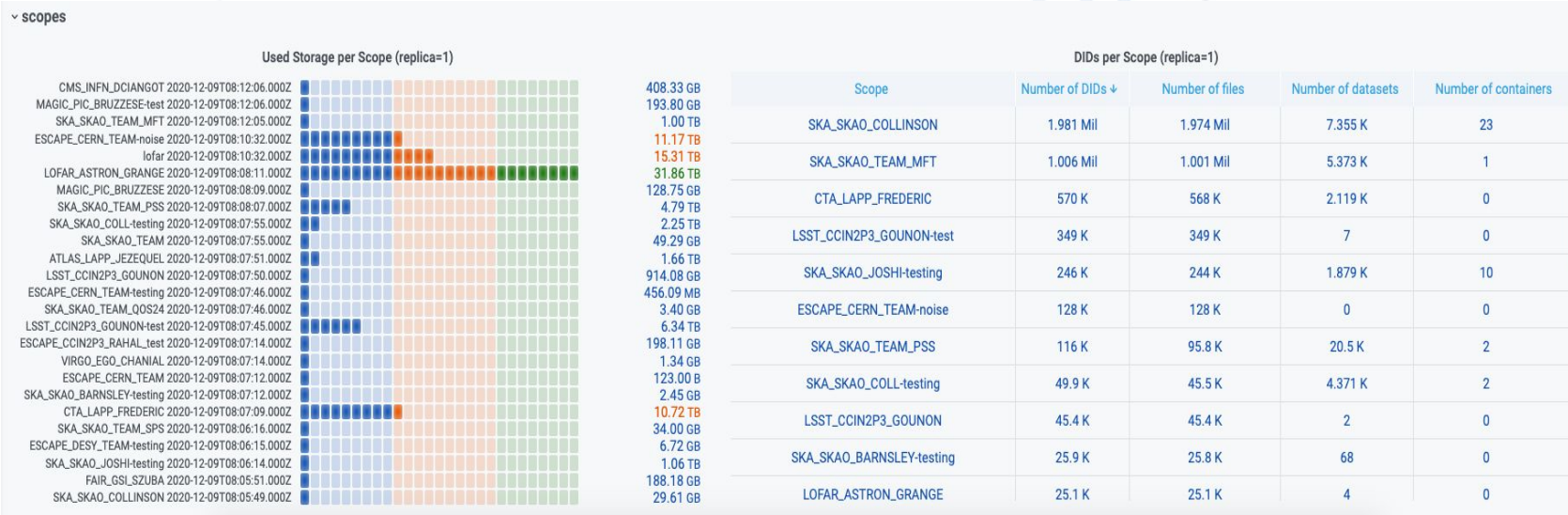
Rucio Dashboards - Uploads (1/2)

- Monitoring replica creation + **Rucio uploads** (an insight Rucio Events cannot provide)

- Custom data pushed by SKAO at their ES

- Based on the SKAO Monitoring setup (hosted on STFC Cloud)

  - ElasticSearch deployed on a 3 node bare-metal kubernetes cluster using the helm charts in this repo

☐	escape-datalake-monitoring-worker-2	Ubuntu-Bionic-NoGui	192.168.200.87	c2.large	rmb-keypair
☐	escape-datalake-monitoring-worker-1	Ubuntu-Bionic-NoGui	192.168.200.31	c2.large	rmb-keypair
☐	escape-datalake-monitoring-head	Ubuntu-Bionic-NoGui	192.168.200.166	c3.large	rmb-keypair

  - Persistent Volume provisioner support using NFS server provisioner

  - Largely a dedicated cluster for monitoring at the moment. It is running ES, and any tools needed to enable ES: NFS, Metal load balancer (MetalLB), Ingress controller

```
rjoshi@escape-datalake-monitoring-head:~$ kubectl get pods -n monit
NAME                                                   READY   STATUS    RESTARTS   AGE
metadata-elasticsearch-coordinating-only-7454647cb5-5zlcm   1/1     Running   0          5d3h
metadata-elasticsearch-coordinating-only-7454647cb5-dtpsm   1/1     Running   0          5d3h
metadata-elasticsearch-data-0                          1/1     Running   0          5d3h
metadata-elasticsearch-data-1                          1/1     Running   0          5d3h
metadata-elasticsearch-master-0                        1/1     Running   0          5d3h
metadata-elasticsearch-master-1                        1/1     Running   0          5d3h
```

rizart.dona@cern.ch

Future Work

- Migration for FTS dashboard from InfluxDB → ES

- Creation of stable perfSONAR dashboards on Grafana

  - Data is already available

  - Experimental dashboards already exist

  - MadDash currently used as an alternative, need to unify monitoring view

- Interconnect data in Rucio Stats dashboard

- Expose kubernetes cluster monitoring that runs the Rucio daemons

rizart.dona@cern.ch

References

- Elasticsearch, https://www.elastic.co/elasticsearch

- InfluxDB, https://www.influxdata.com

- Grafana, https://grafana.com

- Apache ActiveMQ, http://activemq.apache.org

- ESCAPE Grafana Org, https://monit-grafana.cern.ch/d/cHBQ2NjWz/escape-home?orgId=51

Thank you!

Questions?

rizart.dona@cern.ch